(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 11212728 A

(43) Date of publication of application: 06 . 08 . 99

(51) Int. Cl G06F 3/06
G06F 3/06

(21) Application number: 10012457

(22) Date of filing: 26 . 01 . 98

(71) Applicant: HITACHI LTD

(72) Inventor: ISHIKAWA ATSUSHI
MATSUMOTO YOSHIKO
TAKAMOTO KENICHI

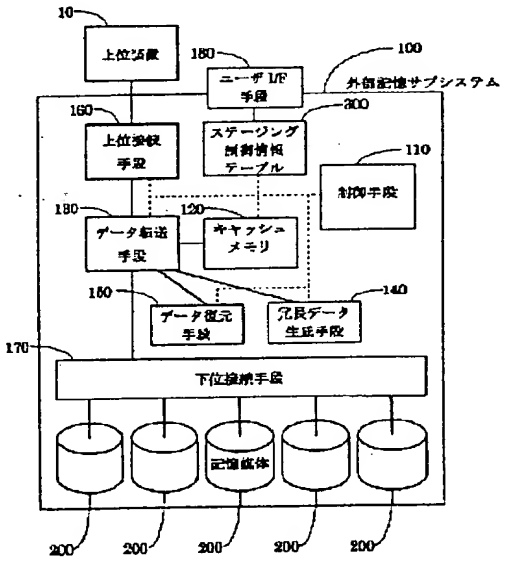
(54) EXTERNAL STORAGE SUB-SYSTEM

COPYRIGHT: (C)1999,JPO

(57) Abstract:

PROBLEM TO BE SOLVED: To provide an external storage sub-system for guaranteeing maximum response time and accelerating single/multiple sequential read.

SOLUTION: A staging control information table 300 for specifying a range for performing staging from a storage medium 200 to a cache memory 120 is provided and a data transfer means 130 performs a staging processing to the cache memory 120 corresponding to the instruction of the staging control information table 300. By simultaneously reading redundant data as well at the time of reading data from the storage medium 200, response time in the case of failing in data read is shortened. Also, since a look-ahead processing is performed without performing look-ahead judgement, the sequential read is accelerated. Further, when a look-ahead direction is specified by the staging control information table 300, since look-ahead in a reverse direction is executed, a high-speed processing is made possible even at the time of the reverse reproduction of the data.



BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平11－212728

(43)公開日　平成11年(1999)８月６日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　3/06	３０２	Ｇ０６Ｆ　3/06	３０２Ａ
	５４０		５４０

審査請求　未請求　請求項の数９　ＯＬ　（全 12 頁）

(21)出願番号　特願平10－12457

(22)出願日　平成10年(1998)１月26日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　石川　篤
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　松本　佳子
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　▲高▼本　賢一
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(74)代理人　弁理士　小川　勝男

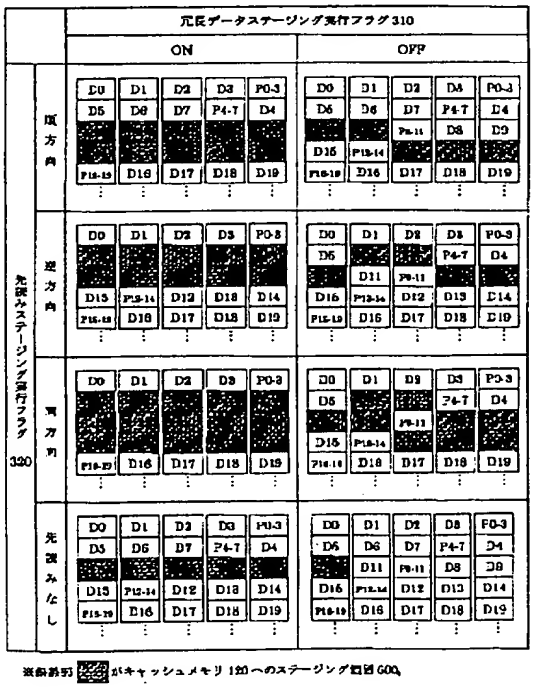
(54)【発明の名称】　外部記憶サブシステム

(57)【要約】

【課題】最大応答時間を保証、および、単一／多重シーケンシャル読み出しの高速化を実現する外部記憶サブシステムを提供する。

【解決手段】記憶媒体からキャッシュメモリにステージングする範囲を指定するステージング制御情報テーブルをもち、ステージング制御情報テーブルの指示にしたがってデータ転送手段がキャッシュメモリにステージング処理をおこなう。記憶媒体からのデータ読み出し時に冗長データも同時に読み出すことにより、データ読み出しに失敗した場合の応答時間を短縮できる。また、先読み判定をおこなわずに先読み処理をおこなうため、シーケンシャルな読み出しを高速化できる。さらに、ステージング情報テーブルによって先読み方向を指定すれば、逆方向への先読みを実施できるので、データの逆再生時にも高速な処理が可能となる。

図6



【特許請求の範囲】

【請求項１】所定の単位に分割した複数のデータを記憶する記憶媒体と、上位装置に接続されて前記記憶媒体とのデータ転送を制御するデータ転送手段と、前記上位装置と前記記憶媒体との間で転送されるデータを一時的に記憶するキャッシュメモリと、複数のデータから冗長データを生成する冗長データ生成手段と、複数のデータと冗長データとから元のデータを復元するデータ復元手段と、前記冗長データと該冗長データを生成した複数のデータとをパリティグループとして管理する制御手段と、前記データを記憶媒体から読み出すときに、該データの冗長データあるいは該データに連続するデータを読み出すかどうかに関する情報を記憶するステージング情報テーブルとを有する外部記憶サブシステム。

【請求項２】前記ステージング情報テーブルは、記憶媒体から読み出すデータの冗長データを読み出すかどうかを決定する冗長データステージング実行フラグを含むことを特徴とする請求項１記載の外部記憶サブシステム。

【請求項３】前記冗長データステージング実行フラグが冗長データのステージングを指示している場合は、記憶媒体からデータを読み出すときに、該データが含まれるパリティグループのデータを読み出すことを特徴とする請求項２記載の外部記憶サブシステム。

【請求項４】前記ステージング情報テーブルは、記憶媒体から読み出すデータに連続するデータを読み出すかどうかを決定する先読みステージング実行フラグを含むことを特徴とする請求項１記載の外部記憶サブシステム。

【請求項５】前記先読みステージング実行フラグが順方向の先読みを指示している場合、記憶媒体からデータを読み出すときに、データの並び順と同じ方向に連続したデータを読み出すことを特徴とする請求項４記載の外部記憶サブシステム。

【請求項６】前記先読みステージング実行フラグが逆方向の先読みを指示している場合、記憶媒体からデータを読み出すときに、データの並び順と反対の方向に連続したデータを読み出すことを特徴とする請求項４記載の外部記憶サブシステム。

【請求項７】前記先読みステージング実行フラグが両方向の先読みを指示している場合、記憶媒体からデータを読み出すときに、データの並び順と同じ方向及び反対の方向に連続したデータを読み出すことを特徴とする請求項４記載の外部記憶サブシステム。

【請求項８】前記ステージング情報テーブルに冗長データステージング実行テーブルあるいは先読みステージングテーブルの内容を入力する手段を有することを特徴とする請求項１記載の外部記憶サブシステム。

【請求項９】前記上位装置から冗長データステージング実行フラグあるいは先読みステージング実行フラグに関する情報を受け取ったとき、前記制御手段は、前記ステージング情報テーブルに冗長データステージング実行フラグあるいは先読みステージング実行フラグの内容を入力する手段を有することを特徴とする請求項１記載の外部記憶サブシステム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、コンピュータ装置に接続される外部記憶サブシステムに係わり、特に応答時間の保証が要求される外部記憶サブシステム、および、多重のシーケンシャルな読みだしにおいて高速なデータ転送能力を要求される外部記憶サブシステムに関する。

【０００２】

【従来の技術】１．応答時間

従来は、ＲＡＩＤなどの冗長データによるデータ復元機能を有する外部記憶サブシステムにおいて、記憶媒体からキャッシュメモリへのデータＤ１０の読み出し（以下、ステージングと称す）に失敗した場合、データ復元に必要な他のデータと冗長データをステージングしてデータを復元した後に、上位装置にデータ転送することにより信頼性を向上させていた。

【０００３】そのため、ステージングが正常終了した場合の応答時間、

ステージング正常終了時の応答時間＝当該データステージング時間＋データ転送時間＋その他処理時間

に比べ、ステージングに失敗した場合には、

ステージング失敗時の応答時間＝当該データステージング時間＋冗長データステージング時間＋データ転送時間＋データ復元時間＋その他処理時間

となり、応答時間が長くなっていた。

【０００４】特開平７－２００１９１号公報には、データ復元の必要性をいち早く認識して、上位装置からの処理を遅延することなしに予備記憶媒体へのデータを復元することを可能とするディスクアレイ装置に関して述べられている。

【０００５】しかし、このディスクアレイ装置では、ステージングに失敗したＩ／Ｏ要求自身の高速化に関しては考慮されていなかった。

【０００６】２．シーケンシャルリード

外部記憶サブシステムでは、シーケンシャルリードを高速化するために、キャッシュメモリを設け先読み処理をおこなうことが一般的となっている。低速な記憶媒体からのデータを上位装置からの要求に先立ってキャッシュメモリに転送しておくこと（先読み処理）によって、上位装置からの読み出し要求を高速なキャッシュメモリからのデータ転送のみで済ませることができ、高速に処理することが可能となる。

【０００７】従来の外部記憶サブシステムでは、一般に単一のシーケンシャルなリード要求に関しては、連続または同一の領域に対して数回の読み出しを受領したことにより上位装置からシーケンシャルなリード要求を受領

していると判断（シーケンシャル判定）した場合に、先読み処理をおこなって処理の高速化を図っていた。また、複数のシーケンシャルリード要求に関しては、シーケンシャル判定に要する情報を複数セット持ち、先読み処理を多重に動かすことにより処理を高速化していた。

【０００８】しかし、これら技術においては、シーケンシャル判定に要する時間短縮や、シーケンシャル判定に要する情報のセット数により発生する先読み処理多重度の制限をなくすための考慮がされていなかった。

【０００９】

【発明が解決しようとする課題】近年の情報化社会にともない、記憶装置の分野においてもマルチメディア対応に対する要求が高まっている。中でも、ビデオオンデマンドシステムやインターネットサーバシステムなどの画像／音声データを扱うシステムにおいては、サーバは記憶装置から大量の画像／音声などの情報を読みだし、単一または複数のクライアント（利用者）に提供する。

【００１０】上記の画像／音声データを扱うシステムでは、データ読みだしの遅れが画像／音声の乱れや遅延の発生につながるため、記憶装置は最大応答時間を保証することが求められる。しかし、ＲＡＩＤなどの冗長データによるデータ復元手段を有する外部記憶サブシステムにおいて記憶媒体からのデータの読み出しに失敗した場合、記憶媒体に再びアクセスして冗長データを読み出してから冗長データからデータを再生して上位装置に転送するため、応答時間が長くなってしまうという問題点があった。

【００１１】また、上記のシステムで扱われる画像／音声のデータは一般に順次性のある（シーケンシャルな）データであることが多く、記憶装置はシーケンシャルな読みだしに対する高速処理も要求される。特に、ビデオオンデマンドシステムでは、画像の再生／逆再生などが行われるため、データの順方向のシーケンシャル読み出しにも、逆方向のシーケンシャル読み出しにも対応が必要である。さらに、現在のネットワーク環境におけるシステム運用を考えると、これらのシステムは多数のクライアントから同時にアクセスされることが多く、多数のシーケンシャルリード処理を平行して高速に処理することが不可欠となっている。

【００１２】従来の外部記憶サブシステムでは、単一のシーケンシャルな読み出しに対しては、連続または同一の領域に複数の読み出しを受領した場合に後続領域に対する先読み処理を開始する事により処理を高速化し、複数多重のシーケンシャルな読み出しに対しては、シーケンシャル判定に用いる情報を複数セット持ち先読み処理を多重に動作させて高速化する方式が用いられることが多かった。しかし、この方式では先読み処理の多重度はシーケンシャル判定に用いる情報のセット数までに制限され制限以上のシーケンシャル読み出しが多重した場合には性能が低下してしまうことや、判定処理によるオーバヘッドで性能が低下してしまうことが問題となっていた。

【００１３】本発明の目的は、記憶媒体からのデータ読み出しに失敗した場合にも短い時間での応答を可能にし、さらに、単一または複数多重の順方向／逆方向のシーケンシャルな読み出しの処理能力を向上させ、シーケンシャルリードの多重度に対する制限を解消する外部記憶サブシステムを提供することである。

【００１４】

【課題を解決するための手段】上記の目的を実現するために、本発明による外部記憶サブシステムでは、データ転送手段により記憶媒体からキャッシュメモリにステージングされる領域をユーザが指定する為に用いるステージング制御情報テーブルを設ける。

【００１５】上位装置からリード要求を受領した場合、データ転送手段はステージング制御情報テーブルのユーザ指示に従って記憶媒体からキャッシュメモリにステージングする範囲を決定し、ステージング処理をおこなう。このステージング制御情報テーブルには、データ復元に必要な冗長データのステージングの有無、および、データＤ１０に連続する領域に対する先読みステージングの有無、先読みの方向（順方向／逆方向／両方向）、先読みをおこなう量を設定する。このステージング制御情報テーブルは、上位装置や内蔵または外部接続されたユーザインタフェース手段を用いて設定／変更が可能であり、ユーザは外部記憶サブシステムを使用するシステムの読み出し要求の特性にあわせてキャッシュメモリへのステージング範囲を変更できる。

【００１６】このステージング制御情報テーブルに、冗長データステージングの実行を指示することによって、上位装置からのリード要求時に、データ転送手段はリード要求のあった当該データとともにその冗長データもキャッシュメモリにステージングする。これにより、当該データの読み出しに失敗した場合でも、ステージング済の冗長データからデータを復元して速やかにデータを送信することができるため、当該データのステージングに失敗した場合の応答時間を改善できる。この制御方式を用いることにより応答時間を

《改善前》：データステージング時間＋冗長データステージング時間＋データ復元時間＋その他時間

から

《改善後》：データステージング時間＋データ復元時間＋その他時間

と冗長データステージング時間の分だけ短縮することができる。

【００１７】また、ステージング制御情報テーブルに先読みステージング処理実行／先読み方向と先読み量を指示すれば、上位装置からの読み出し要求受領時に当該データに順方向／逆方向に連続する領域もキャッシュメモリにステージングする先読みステージング処理がおこな

われる。この方式では、先読み判定をおこなわないため、判定によるオーバヘッドを回避して、シーケンシャルリード処理を高速化できる。また、先読み判定の情報も不要となり先読み判定情報の数による先読み多重度の制限がなくなるため、何多重でも先読みを動かすことが可能となる。また、逆方向への先読みを実施することにより、データの逆再生時にも高速な処理が可能となる。

【００１８】

【発明の実施の形態】以下、本発明の実施形態の１例を図面を用いて説明する。

【００１９】本実施例では５台の記憶媒体にＲＡＩＤ５方式で冗長化してデータを記録する例を示すが、記憶媒体の台数、データ冗長化／復元の方式、データと冗長データの配置などは、任意であり、本実施例以外の方式でもよい。

【００２０】図１は、本発明による外部記憶サブシステムの構成の概略をあらわす図である。外部記憶サブシステム１００は上位装置１０に接続されており、データを格納する記憶媒体２００、装置全体の制御を行う制御手段１１０、上位装置１０とのデータ転送に用いられるキャッシュメモリ１２０、外部記憶サブシステム１００内部のデータ転送や上位装置１０とのデータ転送を制御するデータ転送手段１３０、冗長データを生成する冗長データ生成手段１４０、冗長データからデータを復元するデータ復元手段１５０、上位装置１０とのインタフェースを制御する上位接続手段１６０、下位の記憶媒体とのインタフェースを制御する下位接続手段１７０、記憶媒体２００からキャッシュメモリ１２０へのデータ転送（以下、ステージングとよぶ）の制御に用いるステージング制御情報テーブル３００及びユーザがステージング制御情報テーブル３００などの装置設定を変更するためのインタフェース部として内蔵または接続されているユーザインタフェース手段１８０を有している。

【００２１】この実施例に示した構成はあくまでも１例であり、別の構成による実現も可能である。

【００２２】また、シーケンシャルリードの高速化のみを実現する場合には、冗長データ生成手段１４０やデータ復元手段１５０などのデータ冗長化に関する機能のない構成も可能である。

【００２３】図２は、記憶媒体２００におけるデータと冗長データの配置を示した図である。ここでは、５台の記憶媒体がＲＡＩＤ５方式により冗長化された場合のデータおよび冗長データの配置例を示している。

【００２４】データはストライプと呼ばれる管理単位に分割される。冗長データ生成手段１４０では４つのデータストライプから１つの冗長データストライプを生成する。例えば、データストライプＤ０～Ｄ３から冗長データストライプＰ０－３を生成し、データストライプＤ４～Ｄ７から冗長データストライプＰ４－７を生成する。そしてデータストライプおよび冗長データストライプは５台の記憶媒体２００を巡回するように配置する。このような４つのデータストライプと１つの冗長データストライプの組はパリティグループと呼ばれ、データ復元手段１５０を用いれば、同一のパリティグループの他の３つのデータストライプと１つの冗長データストライプから、残り１つのデータストライプを復元する事が可能である。例えば、データストライプＤ２は、同一のパリティグループＰＧ０－３の３つのデータストライプＤ０／Ｄ１／Ｄ３と冗長データストライプＰ０－３から復元できる。

【００２５】先にも述べたように、記憶媒体の台数、データ冗長化／復元方式、データと冗長データの配置などは、この実施例以外の方法でも構わない。また、本発明によるシーケンシャルリードの高速化のみを実現する場合には、データ冗長化のないデータ配置も可能である。

【００２６】図３では、ステージング制御情報テーブル３００の内容について説明する。

【００２７】ステージング制御情報テーブル３００は、上位装置１０からのリード要求に対しキャッシュメモリ１２０にステージングする範囲を指定する情報を格納するテーブルであり、冗長データステージング実行フラグ３１０、先読みステージング実行フラグ３２０、順方向先読み量３３０、逆方向先読み量３４０からなる。冗長データステージング実行フラグ３１０は、上位装置１０から要求されたデータをステージングする際に当該データＤ１０を復元するときに必要なデータ（本実施例では、同一パリティグループの他のデータと冗長データ）のステージング（以下、冗長データステージングとよぶ）を行うかどうか制御する為に用いられ、ＯＮ／ＯＦＦが設定され、ＯＮの場合には冗長データステージングを実行する指示されているものとみなす。先読みステージング実行フラグ３２０は、上位装置１０から要求されたデータをステージングする際に当該データＤ１０に連続する領域もステージング（以下、先読みステージングとよぶ）の実行と先読みの方向を制御するために用いられ、順方向の先読み、逆方向の先読み、両方向の先読み、先読みなしの４つを指定できるものとする。順方向先読み量３３０には順方向に先読みステージングする先読み量、逆方向先読み量３４０には逆方向に先読みステージングする先読み量を設定でき、キャッシュメモリ１２０の容量を最大値に制限される。本実施例で示した制御情報テーブルは、あくまでも、発明を実現するためのテーブル構成の１例であり、各情報の持ち方・設定値などの違う別のテーブル構成も可能である。

【００２８】ステージング情報管理テーブルは上位装置１０またはユーザインタフェース手段１８０を通して指定できるので、ユーザは使用するシステムに合わせてステージング範囲を指定できる。上位装置１０からステージング制御情報テーブル３００を指定するときは、上位装置からステージング情報管理テーブルの内容に関する

パラメータを有するＳＣＳＩのMODE　SELECTコマンドやベンダユニークのコマンドなどを発行し、コマンドの受領を認識した制御手段によりステージング情報管理テーブル３００を更新する方法があげられる。ユーザインタフェース手段１８０を通してステージング制御情報３００を指定する方法の例としては、ユーザインタフェース手段１８０から上位装置と同様なMODE SELECTコマンドやベンダユニークコマンドを発行する方法や、ユーザインタフェース手段１８０が直接ステージング制御情報テーブル３００を書き換える方法、ユーザインタフェース手段１８０としてディップスイッチを用いてステージング制御情報を設定する方法などがある。これらのステージング制御情報テーブル３００の設定方法も例であり、他の設定手段も可能である。

【００２９】また、本実施例では、簡単のためにステージング制御情報テーブル３００が外部記憶サブシステム１００に１つしかない場合について示したが、外部記憶サブシステム１００内がＲＡＩＤグループなどで領域分割されている場合には分割領域毎にステージング制御情報テーブル３００を用意し、領域毎にステージング方式を変更することも可能である。また、外部記憶サブシステム１００に複数の上位装置１０が接続される場合には、上位装置１０毎にステージング制御情報テーブル３００を用意し、上位装置１０毎にステージング方式を変更することも可能である。

【００３０】さらに、本実施例では、冗長データステージングと先読みステージングの両方を実行可能な例を示したが、冗長ステージングだけあるいは先読みステージングだけが実行可能なテーブル構成も可能である。冗長データステージングのみを可能とする場合、ステージング制御情報テーブル３００は冗長データステージング実行フラグ３１０のみで構成される。先読みステージングのみを可能とする場合、ステージング制御情報テーブル３００は先読みステージング実行フラグ３２０および順方向先読み量３３０、逆方向先読み量３４０のみで構成される。

【００３１】本実施例に上位装置１０からのＩ／Ｏ要求が発行された場合の処理方式の１例を図４～図７を用いて説明する。ここでは、上位装置１０からデータストライプＤ１０にＩ／Ｏ要求が来た場合ついて説明するが、他のデータストライプにＩ／Ｏ要求が来た場合でも同様の処理となる。

【００３２】図４は、本実施例におけるＩ／Ｏ処理の全体の概略を示すフローチャートである。上位装置１０からデータストライプＤ１０のリード要求を受領した（ステップ４０１）場合、制御手段１１０はまず上位装置１０からのＩ／Ｏ要求種別の判定を行い（ステップ４０２）、リード要求であることを認識すると、次にキャッシュメモリ上に当該データＤ１０がステージングされているかどうか確認する（ステップ４０３）。既にキャッシュメモリに当該データＤ１０がステージングされている場合には、ステージング処理など（ステップ４０４～４０８）を行わずに、キャッシュメモリ１２０から上位装置１０に当該データＤ１０を転送し（ステップ４０９）、上位装置１０にＩ／Ｏ処理の完了を報告して（ステップ４１０）処理を完了する。

【００３３】また、キャッシュメモリに当該データＤ１０がステージングされていない場合は、当該データＤ１０をステージング実行中かどうか確認（ステップ４０４）し、当該データＤ１０のステージング処理を実行中ならば実行中のステージング処理が完了するまで待ち（ステップ４０５）、当該データＤ１０のステージング処理を未実行ならばステージング処理を実行する（ステップ４０６）。

【００３４】ステージング処理完了後、当該データＤ１０のステージングが正常に終了したかどうか判定する（ステップ４０７）。ここで、当該データＤ１０のステージング失敗の場合にはデータ復元処理で当該データＤ１０の復元を行う（ステップ４０８）。その後、キャッシュメモリ１２０から上位装置１０に当該データＤ１０を転送し（ステップ４０９）、上位装置１０にＩ／Ｏ処理の完了を報告して（ステップ４１０）処理が完了となる。

【００３５】上位装置１０からライト要求を受領した（ステップ４０１）場合には、Ｉ／Ｏ要求種別判定でライト処理である事を認識（ステップ４０２）したあと、ライトに必要な処理（ステップ４２０）を行った後、上位装置１０にＩ／Ｏ処理の完了を報告する（ステップ４１０）。

【００３６】図５は、ステージング処理（ステップ４０６）を詳細に説明したフローチャートである。ステージング処理（ステップ４０６）では、まず、ステージング制御情報テーブル３００の設定値を元にステージング範囲６００を決定する（ステップ５０１）。冗長データステージング実行フラグ３１０と先読みステージング実行フラグ３２０の組合せとステージング範囲６００の関係については、図６で１例を説明する。

【００３７】次に、ステージング範囲６００のデータストライプ分だけループして（ステップ５０２）、キャッシュ上に書くデータストライプがステージングされているかどうか確認し（ステップ５０３）、キャッシュ上にデータストライプがステージングされていない場合には、記憶媒体２００にデータストライプの読み出し要求を発行する。

【００３８】図６に、冗長データステージング実行フラグ３１０及び先読みステージング実行フラグ３２０の設定値とステージングされる領域の関係をまとめる。

【００３９】本実施例では、冗長データステージング実行フラグ３１０の設定値はＯＮ／ＯＦＦ、先読みステージング実行フラグ３２０の設定値は順方向、逆方向、両

方向、先読みなしの４種類が設定でき、順方向先読み量３３０および逆方向先読み量３４０にはそれぞれ４データストライプ分の先読みが指示されている場合を取り上げたが、テーブル構造や設定値が異なる場合でも同様の方法でステージング範囲６００を決定する。また、上位装置１０からデータストライプＤ１０にリード要求が発行された場合であるが、他のデータストライプにリード要求が発行された場合に関しても同様である。また、本実施例では、順方向先読み量３３０に４データストライプ分の先読みを実行されるように指示されているものとして説明しているが、本実施例以外の方法による順方向先読み量３３０の設定も可能である。

【００４０】冗長データステージング実行フラグ３１０がＯＮ、先読みステージング実行フラグ３２０が順方向と設定されている場合、ステージング範囲６００は当該データＤ１０を含むパリティグループＰＧ８－１１とそれに順方向に連続するパリティグループＰＧ１２－１５となる。

【００４１】冗長データステージング実行フラグ３１０がＯＮ、先読みステージング実行フラグ３２０が逆方向と設定されている場合、ステージング範囲６００は当該データＤ１０を含むパリティグループＰＧ８－１１とそれに逆方向に連続するパリティグループＰＧ４－７となる。

【００４２】冗長データステージング実行フラグ３１０がＯＮ、先読みステージング実行フラグ３２０が両方向と設定されている場合、ステージング範囲６００は当該データＤ１０を含むパリティグループＰＧ８－１１とそれに順方向／逆方向に連続するパリティグループＰＧ１２－１５／ＰＧ４－７となる。

【００４３】冗長データステージング実行フラグ３１０がＯＮ、先読みステージング実行フラグ３２０が先読みなしと設定されている場合、ステージング範囲６００は当該データＤ１０を含むパリティグループＰＧ８－１１のみとなる。

【００４４】冗長データステージング実行フラグ３１０がＯＦＦ、先読みステージング実行フラグ３２０が順方向と設定されている場合、ステージング範囲６００は当該データＤ１０とそれに順方向に連続する順方向先読み量３３０分のデータストライプＤ１１～Ｄ１４となる。

【００４５】冗長データステージング実行フラグ３１０がＯＦＦ、先読みステージング実行フラグ３２０が逆方向と設定されている場合、ステージング範囲６００は当該データＤ１０とそれに逆方向に連続する逆方向先読み量３４０分のデータストライプＤ９～Ｄ６となる。

【００４６】冗長データステージング実行フラグ３１０がＯＦＦ、先読みステージング実行フラグ３２０が両方向と設定されている場合、ステージング範囲６００は当該データＤ１０とそれに順方向に連続する順方向先読み量３３０分のデータストライプＤ１１～Ｄ１４、および、逆方向に連続する逆方向先読み量３４０分のデータストライプＤ９～Ｄ６となる。

【００４７】冗長データステージング実行フラグ３１０がＯＦＦ、先読みステージング実行フラグ３２０が先読みなしと設定されている場合、ステージング範囲６００は当該データＤ１０のみとなる。

【００４８】図６の中では、各組み合わせにおいて斜線で示したデータストライプがステージング範囲６００である。

【００４９】図７は、データ復元処理（ステップ４０８）を説明したフローチャートである。

【００５０】ステージング処理の終了判定（ステップ４０７）において、当該リードデータストライプＤ１０の読み出しに失敗した場合、冗長データから当該データＤ１０を復元するデータ復元処理（ステップ４０８）が必要となる。

【００５１】データ復元処理（ステップ４０８）では、まず、冗長データステージング処理の実行状態を確認する（ステップ７０１）。ここで、冗長データステージング処理を未実行の場合には冗長データステージング処理を実行し、当該データＤ１０を復元するのに必要な同一パリティグループＰＧ８－１１の他のデータストライプＤ８／Ｄ９／Ｄ１１と冗長データストライプＰ８－１１がキャッシュメモリ１２０にステージングする（ステップ７０２）。また、冗長データステージング処理が実行中でまだ終了していない場合には冗長データステージング処理の終了を待つ（ステップ７０３）。冗長データのステージングが完了した後、データ復元手段１５０を用いて当該データＤ１０を復元してキャッシュメモリ１２０に格納する（ステップ７０４）。

【００５２】本実施例においてユーザが冗長データステージング実行を指示した場合の応答時間の短縮効果について説明する。

【００５３】従来方式では当該データＤ１０のステージング実行時に冗長データステージングを実行していない。そのため、当該データＤ１０のステージング処理失敗が判明した後、改めてデータ復元に必要な同一パリティグループＰＧ８－１１の他のデータＤ８／Ｄ９／Ｄ１１と冗長データＰ８－１１のステージングを実行してから、データ復元をおこなっていた。すなわち、従来方式における当該データＤ１０のステージングに失敗した場合の応答時間は、

従来方式での応答時間＝当該データステージング時間＋冗長データステージング時間＋データ復元時間＋その他時間

となっていた。

【００５４】これに対し、本発明の方式で、ユーザが上位装置１０やユーザインタフェース手段１８０を用いて冗長データステージング実行フラグ３１０にＯＮを設定した場合には、当該データＤ１０ステージング時に同時

に冗長データステージングも行う（ステップ５０３、または、ステップ５０４）ため、データ復元処理（ステップ４０８）時には冗長データステージングが不要となり、キャッシュメモリ１２０上の冗長データを用いて速やかにデータ復元が行える。そのため、応答時間は、
本発明の方式での応答時間＝当該データステージング時間＋データ復元時間＋その他時間
となり、従来方式に比べて、冗長データステージング時間の分だけ応答時間を短縮する事ができる。一般に、記憶媒体２００からのステージングは、キャッシュメモリ１２０からの読み出しに比べて非常に遅く、本発明により応答時間を大きく短縮する事ができる。

【００５５】次に、本実施例においてユーザが先読みステージング実行を指示した場合のシーケンシャルリード処理の高速化に対する効果について説明する。

【００５６】ユーザが上位装置１０やユーザインタフェース手段１８０を用いて先読みデータステージング実行フラグ３２０に順方向を設定した場合には、当該データＤ１０をステージングする時にＤ１０に順方向に連続する領域（Ｄ１１以降）のデータも同時にステージングする。また、先読みデータステージング実行フラグ３２０に逆方向を設定した場合には、当該データＤ１０をステージングする時にＤ１０に逆方向に連続する領域（Ｄ９以前）のデータも同時にステージングし、先読みデータステージング実行フラグ３２０に両方向を設定した場合には、当該データＤ１０をステージングする時にＤ１０に順方向／逆方向に連続する領域（Ｄ１１以降、および、Ｄ９以前）のデータも同時にステージングする。

【００５７】このとき、従来行っていたシーケンシャル判定は不要となるため、従来方式のシーケンシャル判定に要していた時間が削減できる。

【００５８】また、上位装置１０から順方向／逆方向のシーケンシャルなリード要求が発行される場合、１回目のリード要求時に２回目以降のリード要求の領域も先にステージングされているため、２回目以降のリード処理ではキャッシュメモリからホストへデータ転送するだけでよいので処理時間を短縮することができる。一般に、記憶媒体２００からのステージングにくらべてキャッシュメモリ１２０からの読み出しは非常に早いため、本発明によりシーケンシャルリード処理を大きく高速化できる。

【００５９】最後に、本実施例においてユーザが先読みステージング実行を指示した場合の複数多重シーケンシャルリード処理方式、および、効果について説明する。

【００６０】本発明では、シーケンシャルリード要求の多重数により制限される情報を用いないため、シーケンシャルリード要求の多重度を意識することなく処理が可能である。複数多重シーケンシャルリード要求の場合でも、上位装置１０からのリード要求時にステージング制御情報テーブル３００の設定内容にしたがって記憶媒体２００へステージング要求を発行する。ここで、下位へのＩ／Ｏ要求を管理するために一般的に使用されている、キューイング機能を持つ記憶媒体２００やキューイング機能を持つ下位接続手段１７０を使用することにより、多重に発行したステージング要求を管理／実行することが可能となる。

【００６１】したがって、従来は一般にシーケンシャル判定により先読みステージングを行っているためシーケンシャル判定を行うためのテーブルのセット数により先読みステージング処理の多重度が制限されていたが、本発明ではシーケンシャル判定が不要なためシーケンシャルリード処理の多重度に関する制限がなくなり、何多重でも先読みステージング処理を実行できるため、多数多重したシーケンシャル要求に対しても高速に処理することができる。

【００６２】本発明によれば、上位装置１０やユーザインタフェース手段１８０を通して設定することによって、従来方式のステージング、冗長データも含めたステージング、順方向／逆方向の先読みステージングの中から、使用するシステムにあわせたステージング方式を選択しチューニングすることができる。また、ＲＡＩＤグループやＬＵなどで外部記憶サブシステム１００が領域分割されている場合や、複数ホストが接続されている場合には、その領域分割単位やホスト単位にステージング制御情報テーブル３００を持つことによって、領域分割毎、ホスト毎のステージング方法のチューニングも可能となる。

【００６３】

【発明の効果】本発明によれば、冗長データによるデータ復元手段を持つ外部記憶サブシステムにおいて記憶媒体からのデータ読出しに失敗した時に速やかにデータ復元をおこなえるため、応答時間を短くでき最大応答時間保証に効果がある。

【００６４】また、先読みステージング処理実行／先読み方向などの先読みに関する情報を保持することにより、上位装置からのシーケンシャルなデータ読み出しに対してシーケンシャル判定が不要になるため、処理時間を短縮でき、単一および複数多重のシーケンシャルな読み出しの処理高速化に効果がある。

【００６５】さらに、シーケンシャル判定情報を保存しておく必要がなくなるため、シーケンシャル判定情報の数による先読み多重数の制限がなくなる。したがって、何多重にも先読み処理を動作させることができるため、多重シーケンシャル読み出しの高速化に効果がある。また、逆方向への先読みを実施することにより、データの逆再生時にも高速な処理が可能となる。

【図面の簡単な説明】

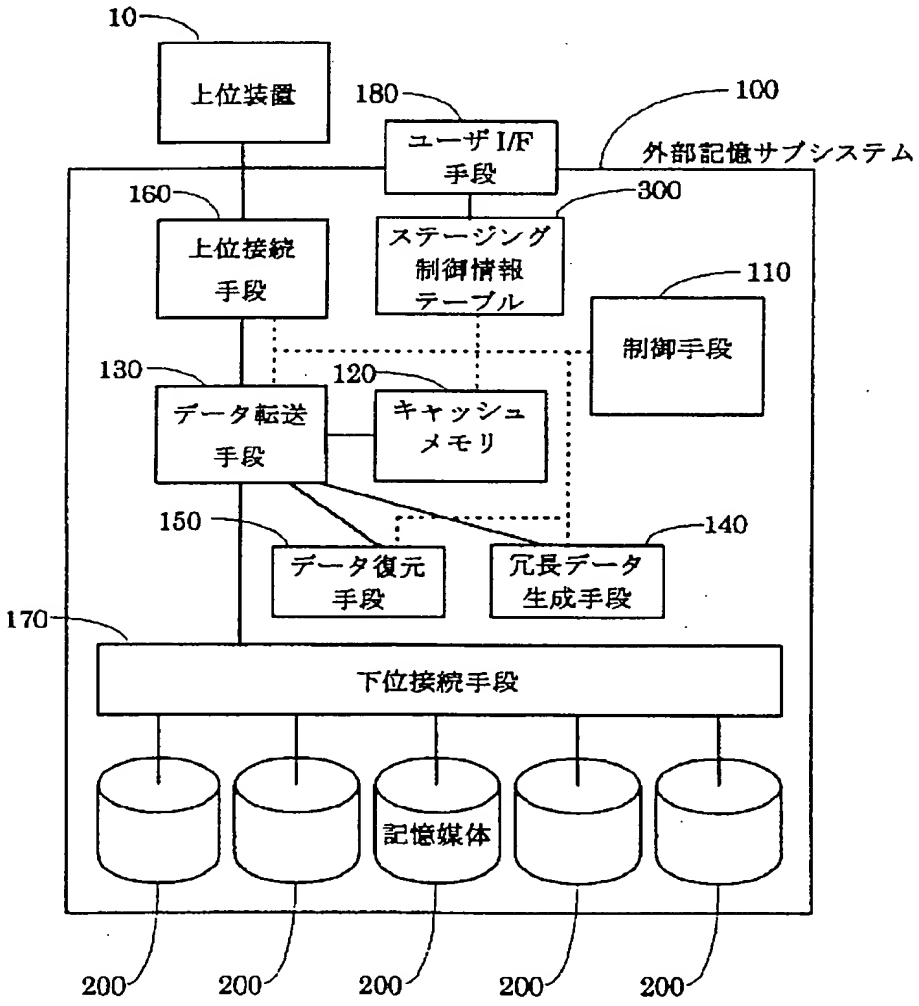
【図１】本発明による外部記憶サブシステムの構成を示す図である。

【図２】外部記憶サブシステムの記憶媒体内部のデータ

／冗長データの配置の１例を示す概念図である。
【図３】本発明による外部記憶サブシステムで用いられるステージング制御情報テーブルの例を示す図である。
【図４】本発明による外部記憶サブシステムのＩ／Ｏ処理のフローチャートである。
【図５】ステージング処理を詳細に示すフローチャートである。
【図６】ステージング制御情報テーブルの設定値とステージング実行範囲との関係の例を示す概念図である。
【図７】データ復元処理を示すフローチャートである。
【符号の説明】
１０：上位装置
１００：外部記憶サブシステム
１１０：制御手段　　１２０：キャッシュメモリ
１３０：データ転送手段　　１４０：冗長データ生成手段
１５０：データ復元手段　　１６０：上位接続手段
１７０：下位接続手段　　１８０：ユーザインタフェース手段
２００：記憶媒体
３００：ステージング制御情報テーブル
３１０：冗長データステージング実行フラグ
３２０：先読みステージング実行フラグ
３３０：順方向先読み量　　３４０：逆方向先読み量

【図１】

図１



【図２】

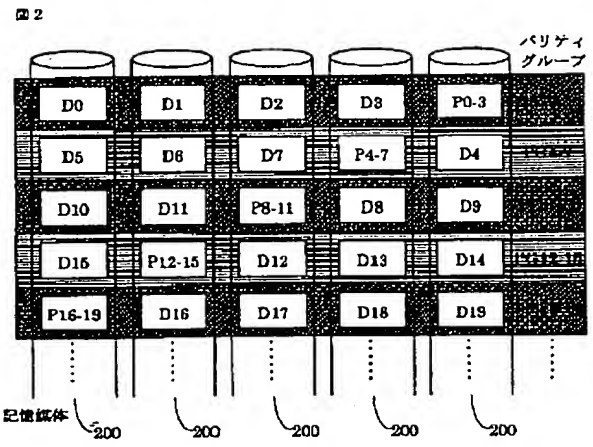
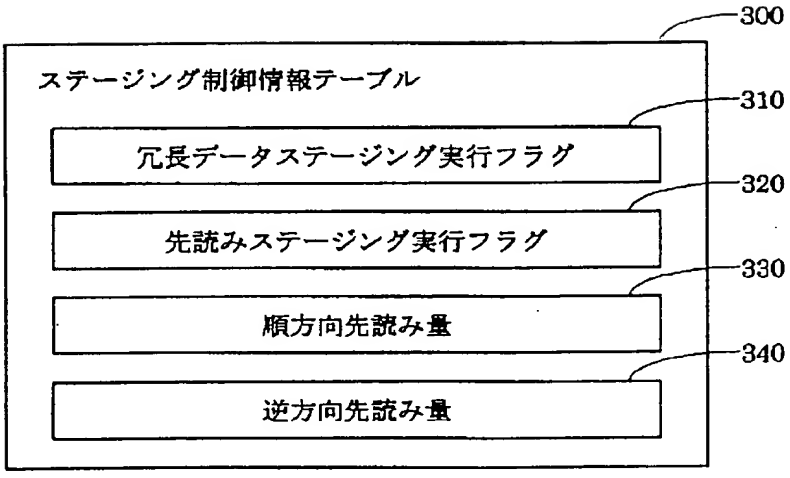


図2
パリティグループ
D0
D1
D2
D3
P0-3
D5
D6
D7
P4-7
D4
D10
D11
P8-11
D8
D9
D15
P12-15
D12
D13
D14
P16-19
D16
D17
D18
D19
記憶媒体
200
200
200
200
200

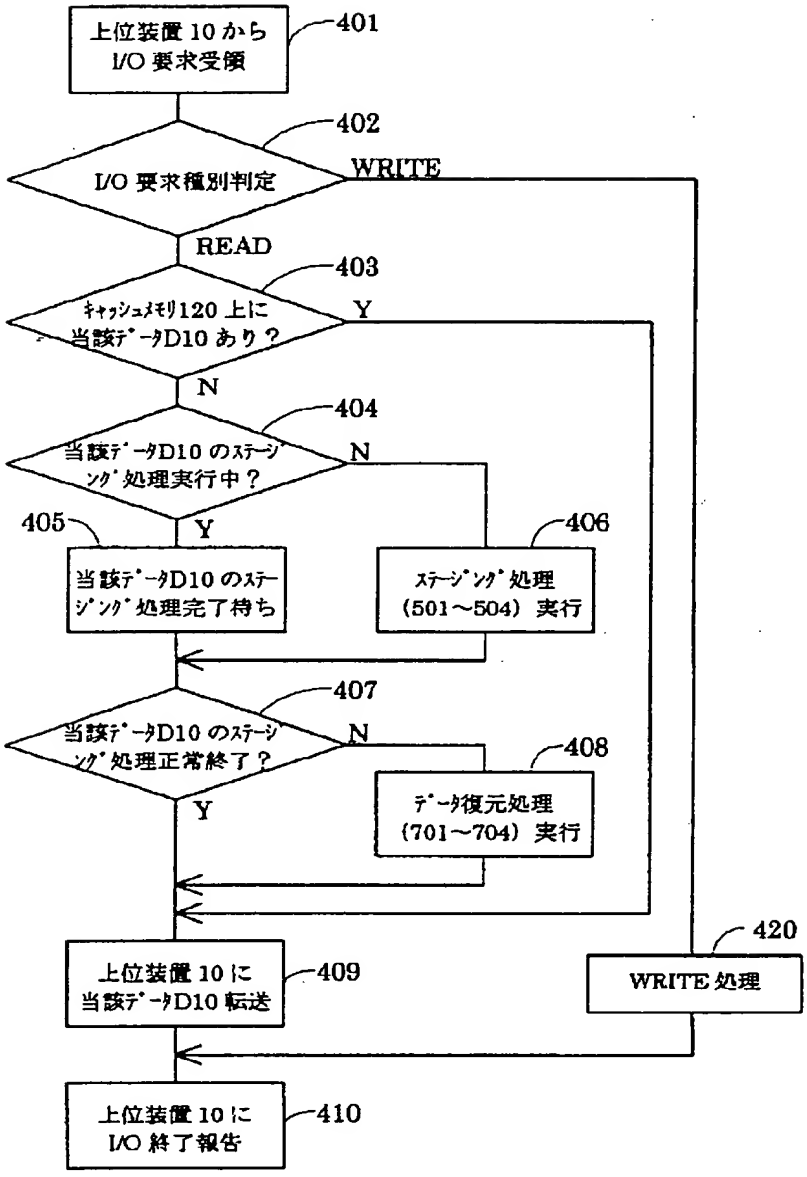
【図３】

図３


300
ステージング制御情報テーブル
310
冗長データステージング実行フラグ
320
先読みステージング実行フラグ
330
順方向先読み量
340
逆方向先読み量

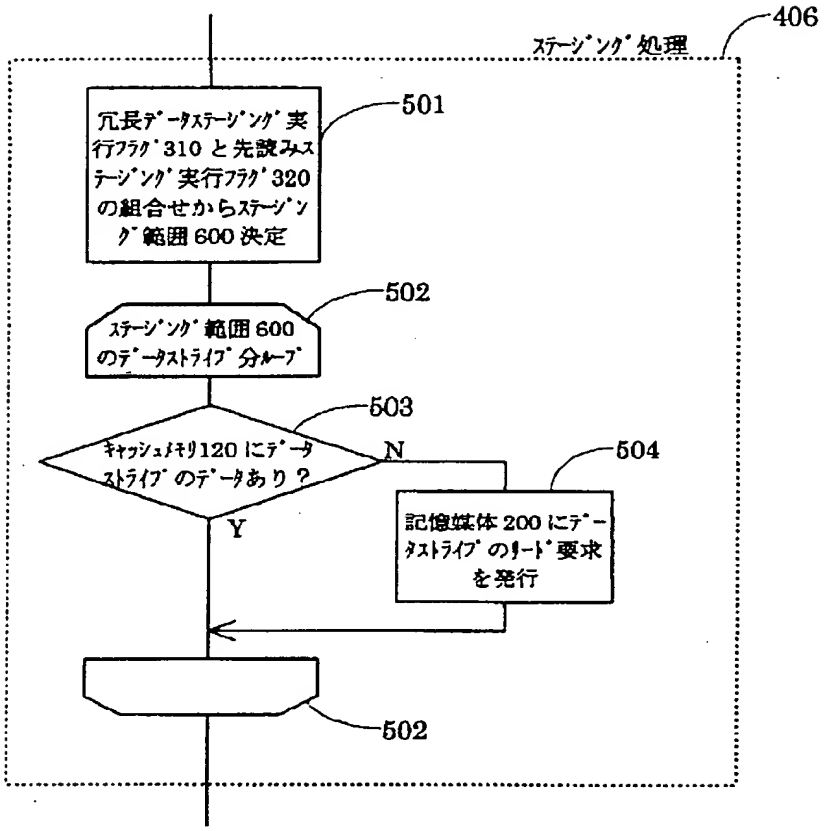
【図４】

図４


上位装置10から
I/O要求受領
401
402
I/O要求種別判定
WRITE
READ
403
ｷｬｯｼｭﾒﾓﾘ120上に
当該ﾃﾞｰﾀD10あり？
Y
N
404
当該ﾃﾞｰﾀD10のｽﾃｰｼﾞ
ﾝｸﾞ処理実行中？
N
Y
405
406
当該ﾃﾞｰﾀD10のｽﾃｰ
ｼﾞﾝｸﾞ処理完了待ち
ｽﾃｰｼﾞﾝｸﾞ処理
(501～504) 実行
407
当該ﾃﾞｰﾀD10のｽﾃｰｼ
ﾝｸﾞ処理正常終了？
N
Y
408
ﾃﾞｰﾀ復元処理
(701～704) 実行
420
上位装置10に
当該ﾃﾞｰﾀD10転送
409
WRITE処理
上位装置10に
I/O終了報告
410

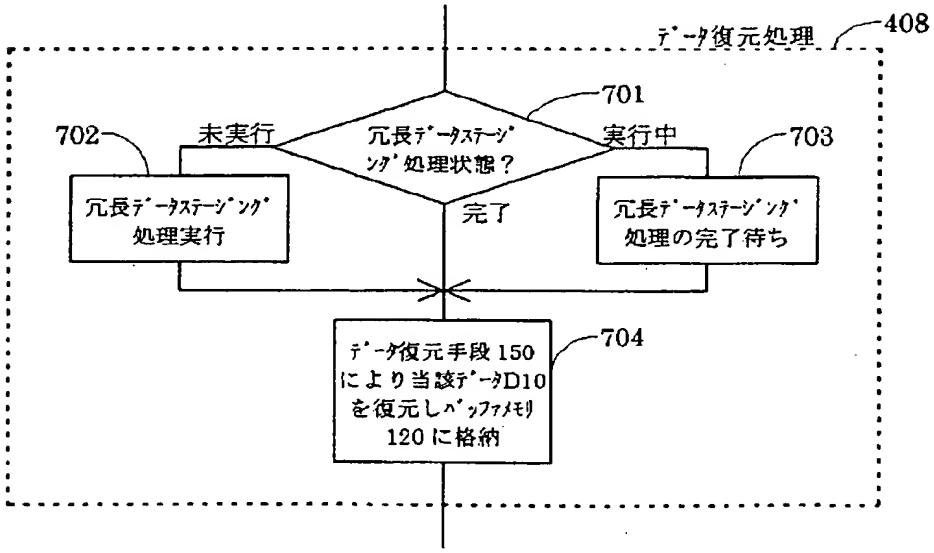
【図５】

図５


406
ステージング処理
501
冗長データステージング実行フラグ310と先読みステージング実行フラグ320の組合せからステージング範囲600決定
502
ステージング範囲600のデータストライプ分ループ
503
キャッシュメモリ120にデータストライプのデータあり？
N
Y
504
記憶媒体200にデータストライプのリード要求を発行
502

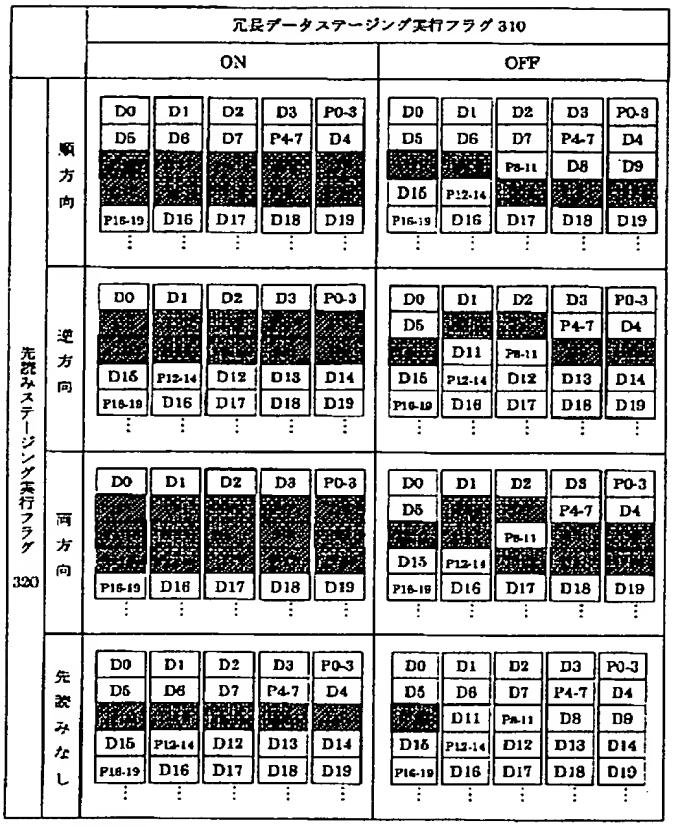
【図７】

図７


408
データ復元処理
701
冗長データステージング処理状態？
未実行
実行中
完了
702
冗長データステージング処理実行
703
冗長データステージング処理の完了待ち
704
データ復元手段150により当該データD10を復元しバッファメモリ120に格納

【図６】

図6



送料部 ▒▒ がキャッシュメモリ 120 へのステージング範囲 600。